FoodMicrobionet: A database for the visualisation and exploration of food bacterial communities based on network analysis

FoodMicrobionet: a database for the visualisation and exploration of food bacterial communities based on network analysis

Eugenio Parente[1], Luca Cocolin[2], Francesca De Filippis[3], Teresa Zotta[4], Ilario Ferrocino[2], Orla O'Sullivan[5], Erasmo Neviani[6], Maria De Angelis[7], Paul D. Cotter[5], Danilo Ercolini[3]

Affiliations

1 Dipartimento di Scienze, Università degli Studi della Basilicata, Potenza, Italy

2 Department of Agricultural, Forest and Food Science, University of Torino, Grugliasco, Italy

3 Department of Agricultural Sciences, Division of Microbiology, University of Naples Federico II, Portici, Italy

4 Istituto di Scienze dell'Alimentazione, CNR, Avellino, Italia

5 Teagasc Food Research Centre, Moorepark, Fermoy and APC Microbiome Institute, Cork, Ireland

6 Department of Food Science, Parma University, Parco Area delle Scienze 48/A, Parma, Italy

7 Department of Soil, Plant and Food Science, University of Bari Aldo Moro, Bari, Italy

Corresponding author

Prof. Eugenio Parente

Dipartimento di Scienze, Università degli Studi della Basilicata, Viale dell'Ateneo Lucano, 10, 85100 Potenza, Italy

E-mail eugenio.parente@unibas.it

tel +390971205561

Abstract

Amplicon targeted high-throughput sequencing has become a popular tool for the culture-independent analysis of microbial communities. Although the data obtained with this approach are portable and the number of sequences available in public databases is increasing, no tool has been developed yet for the analysis and presentation of data obtained in different studies. This work describes an approach for the development of a database for the rapid exploration and analysis of data on food microbial communities. Data from seventeen studies investigating the structure of bacterial communities in dairy, meat, sourdough and fermented vegetable products, obtained by 16S rRNA gene targeted high-throughput sequencing, were collated and analysed using Gephi, a network analysis software. The resulting database, which we named FoodMicrobionet, was used to analyse nodes and network properties and to build an interactive web-based visualisation. The latter allows the visual exploration of the relationships between Operational Taxonomic Units (OTU) and samples and the identification of core- and sample- specific bacterial communities. It also provides additional search tools and hyperlinks for the rapid selection of food groups and OTUs and for rapid access to external resources (NCBI taxonomy, digital versions of the original articles). Microbial interaction network analysis was carried out using CoNet on datasets extracted from FoodMicrobionet: the complexity of interaction networks was much lower than that found for other bacterial communities (human microbiome, soil and other environments). This may reflect both a bias in the dataset (which was dominated by fermented foods and starter cultures) and the lower complexity of food bacterial communities.

Although some technical challenges exist, and are discussed here, the net result is a valuable tool for the exploration of food bacterial communities by the scientific community and food industry.

Key words Food bacterial communities; Network analysis; 16S rRNA amplicon-based high-throughput sequencing.

1. Introduction

The degree of complexity of the microbiota that potentially impacts on food quality and safety is extremely variable. In both fermented and non-fermented foods, the type of contaminating microbiota can initially be rather diverse, reflecting mainly the original microbiota of the raw material, processing, handling and storage conditions and the level of good manufacture practice (Bokulich et al., 2012; Bokulich and Mills, 2013; Chaillou et al., 2015, Cocolin and Ercolini, 2015; De Filippis et al., 2013). However, depending on the storage conditions and other extrinsic factors, only a few species and strains will be able to develop sufficiently in the food matrix to significantly affect the food quality (by spoilage or fermentation) or safety. The methods employed to study microbes and microbial diversity in foods have evolved, and in turn, have also revolutionized our overall understanding of the microbial ecology of foods (Cocolin and Ercolini, 2015). The culture-independent evaluation of food microbial diversity by high-throughput rRNA gene sequencing has become an increasingly popular approach to food microbiology. After microbial nucleic acid extraction from the food, the DNA (or cDNA in cases where RNA was targeted) is used as a template to amplify variable regions within or across the rRNA genes of bacteria (16S) or fungi (internal transcribed spacer [ITS] or other target) and an amplicon library is then sequenced using high-throughput sequencing (HTS; Ercolini, 2013) platforms. The result is a food (sample)-specific profile of the microbiota where all the microbial entities are identified at variable taxonomic depth. Based on the number of sequence reads assigned to a given taxon, the relative abundance of each identified operational taxonomic unit (OTU) can be determined. Therefore, for each food sample

77 analysed, a clear understanding of the composition and relative abundances of the
78 microorganisms populating the food at that time can be provided. The advantages and
79 disadvantages of this methodology have been discussed elsewhere (Bokulich and Mills, 2013;
80 Ercolini, 2013).
81 Many different sequencing technologies are available for the generation of sequence data
82 (Glenn, 2011. http://www.molecularecologist.com/next-gen-fieldguide-2014). Regardless of
83 how the data is generated, accurate data analysis tools are pivotal in any study of microbial
84 ecology; from quality filtering to graphical representations, the software and the algorithms
85 selected can greatly impact on the results and interpretation. Essential steps in any analysis
86 pipeline include post-sequencing quality checking (based on both length and quality scores),
87 clustering into OTUs, chimera removal, alignment, taxonomical assignment and diversity
88 analysis. The choice of the pipeline has been proven to significantly affect the results in terms
89 of estimated diversity and microbial community structure (May et al., 2014). Diversity can be
90 calculated from both a within sample (alpha diversity) and between sample perspective (beta
91 diversity). Numerous packages have been developed for rRNA gene amplicon data analysis,
92 primarily designed for UNIX based operating systems. The most widely used packages are
93 QIIME (Caporaso et al, 2010) and MOTHUR (Schloss et al., 2009). They have become popular
94 because they provide a pre-compiled and user-friendly analysis pipeline, but also due to their
95 constant maintenance and updates.
96 Beyond the power linked to the sensitivity and the throughput of sequencing-based
97 microbiota analysis, a fundamental advantage is the possibility of using the raw sequence data
98 in meta-studies. In fact, in contrast to previous culture-independent approaches, the
99 sequencing-based tools offer the unprecedented advantage of making the results readily
100 available for the scientific community through the deposit of the sequences in public
101 databases (e.g. the Sequence Read Archive (SRA) of the National Center for Biotechnology

Information (http://www.ncbi.nlm.nih.gov/Traces/sra) or the European Nucleotide Archive of the European Bioinformatics Institute (http://www.ebi.ac.uk/ena). This allows researchers to easily access datasets corresponding to diverse food samples generated by different laboratories and with different scopes.

Network analysis (Newman et al., 2006) tools have recently been used to provide effective and information dense displays of microbial communities for several environments (de Menezes et al., 2014; Deng et al., 2012; Muegge et al., 2011; Zhou et al., 2011a), including foods (Chaillou et al., 2015; De Filippis et al., 2013; De Filippis et al., 2014; Dolci et al., 2014; Ercolini et al., 2013; Oakley et al., 2013). In a network representation, objects (OTUs and/or samples) represent the nodes (or vertices), and are connected by links (edges). The edges can be directed (i.e. when the direction of the connection is of importance) or undirected and are usually associated with a weight. The latter can store information on the abundance of an OTU in a sample or the probability of a significant co-occurrence/co-exclusion relationship.

Two types of displays have been used in microbial ecology. In OTU - sample nets, the network is bipartite i.e. two types of nodes exist, sample and OTU nodes, and connections occur only between samples and associated OTUs. Conversely, co-occurrence/co-exclusion networks, which show significant positive or negative interactions among members of microbial communities have rarely (Chaillou et al., 2015; Mounier et al., 2008; Oakley et al., 2013) been used in food microbial ecology, but have been successfully applied to the study of the microbial communities of a variety of environments (miscellaneous environments: Deng et al., 2012; water: Liu et al., 2014; soil: Zhou et al., 2011a) and for the human microbiome (Faust et al., 2012). The methods and models used to derive interaction networks have been reviewed by Faust and Raes (2012) and, although they are riddled by pitfalls related to the structure of the data and to the sensitivity to the methods and parameters selected in the analysis (Faust and Raes, 2012; Kuczynski et al., 2010), they offer significant advantages with respect to

detecting biologically and ecologically relevant relationships among members of microbial communities.

The number of HTS studies of food microbial communities has been increasing steadily in recent years (Mayo et al., 2014), but the information is dispersed in a large number of papers, each analysing a single or a limited range of foods. It is therefore tempting and timely to collect and integrate data from several studies in such a way that the results can be readily searched and visualized, even by relatively inexperienced users. With the aim of providing flexible means for meta-studies in food microbial ecology, here we present FoodMicrobionet, a database and visualisation tool based on network analysis, and some examples of the potential of the tool in terms of data display and analysis.

2. Material and methods

2.1. Data sources

FoodMicrobionet 1.0 includes data from 17 studies, on dairy products, dairy starter cultures, raw and fermented meat, doughs and sourdoughs, or fermented vegetables. The list of studies, with information on the sequencing platforms, software employed for bioinformatics analysis, and the databases used for OTU assignment is shown in Table 1.

2.2 Data tables

Abundance tables, including taxonomic lineages for each OTU, were obtained from each contributor and transformed in tab-delimited nodes and edges tables, which were collated and curated to remove duplicates. The node and edge tables and their specifications are provided in section 1 of Supplementary Material.

2.3. Network analysis

2.3.1. OTU - Food network

151 The edges tables were imported in Gephi 0.8.2-beta (http://gephi.github.io/; Bastian and
152 Jacomy, 2009) using the "Import spreadsheet" feature. Nodes tables were then imported to
153 retrieve the metadata for each node. Statistics (degree and weighted degree, centrality
154 statistics, network diameter, graph density, average path length) were then calculated for
155 each node and for the network using the statistical module of Gephi. A glossary of terms for
156 node and network statistics is provided in Table 2. Styles were then applied to the nodes to
157 enhance the display: the colour of the node was attributed on the basis of a custom field
158 containing families for OTUs and Food subgroup for samples; the size of the nodes was made
159 proportional to the weighted degree of the node; edge thickness was made proportional to the
160 weight of the connection. A Yfan Hu force based layout algorithm was finally applied (Hu,
161 2006). Simplified versions of the networks were obtained by filtering. The whole network was
162 then exported for web visualisation using the Sigmajs exporter plugin of Gephi.
163 2.3.2. Microbial interaction networks
164 Microbial interaction networks were generated for selected groups of samples extracted from
165 FoodMicrobionet using the CoNet app (Faust et al., 2012) of Cytoscape 3.2.1. OTU abundance
166 (as number of sequences per sample) tables were then imported using CoNet, and five
167 methods (Pearson, Spearman, Mutual information, Bray Curtis, Kullback-Leibler) were used
168 to mine for significant co-occurrence/co-exclusion relationships. Null distributions were
169 generated using the edge-scores routine and random distributions using the bootstrap
170 routine. Brown's method was used to merge method specific p-values and the Benjamini
171 Hochberg method was used to adjust the p-values for multiple testing. Interactions were
172 evaluated for both low level taxa and high level taxa, but a parent child exclusion filter was
173 used to avoid interactions between a high level taxon and its members (e.g.. between
174 *Lactobacillaceae* and members of the genus *Lactobacillus*). To simplify network visualisations,
175 interactions with high level taxa were included only if they did not duplicate interaction due

176 to a lower level taxon (i.e. if *Lactobacillales* and *Lactobacillus delbrueckii* shared the same
177 interactions, the former node was removed). Topological properties of the interaction
178 networks were evaluated using the NetworkAnalyzer tool of Cytoscape.
179
180 **3. Results and discussion**
181 *3.1. Building FoodMicrobionet: a bipartite OTU-sample network for food bacterial communities.*
182 The main purpose of FoodMicrobionet is to provide a user-friendly tool to explore multiple
183 datasets generated by 16S rRNA gene amplicon HTS studies of food bacterial communities.
184 The flowchart for the development of FoodMicrobionet and of its products (visualisations,
185 tables, graphs) is shown in Fig. 1. Data from seventeen published and unpublished studies
186 (Table 1) on dairy and meat products, starter cultures, sourdoughs or fermented vegetable
187 products (olives) were assembled in a database and a network was generated using Gephi
188 0.8.2-beta. The network has 964 OTU nodes and 552 sample nodes, with 18,115 edges
189 (sample-OTU relationships), and is by far the largest such collection of data of food bacterial
190 microbiota.
191 Network analysis software packages such as Gephi or Cytoscape (edges and nodes files
192 provided in supplementary material can be easily imported in Cytoscape) offer a wide range
193 of filtering, statistical analysis and graphical representation options but require some
194 informatics skills. Therefore an interactive network visualisation
195 (http://www.foodmicrobionet.org/fmbn1_0_3web/) was created using a publicly available
196 plugin. This visualisation allows even inexperienced users to explore FoodMicrobionet, to
197 select individual sample or OTU nodes, or to carry out group selections for sample and OTU
198 nodes. Relevant properties for both OTU and food sample nodes can be visualised by either
199 clicking on nodes or by selecting them using a search field. A user manual for the web
200 visualisation is provided in section 2 of Supplementary material.

201 Because of the high number of sample and OTU nodes, the information cannot be easily
202 presented into a readable graph. Therefore, simplified, filtered and node-labelled sub-
203 networks for meats, sourdoughs and dairy foods are shown in Fig. 2. The common features
204 found in OTU-sample networks previously published (De Filippis et al., 2013; De Filippis et al.,
205 2014; Dolci et al., 2014; Ercolini et al., 2013) are evident: sample nodes with similar
206 microbiota occupy defined areas of the graph and are close to the OTU nodes that dominate
207 their microbiota. This allows to identify easily the dominant, core and minor OTUs, that can be
208 clearly distinguished by their position and by their node size.
209 Taxon specific sub-networks can be easily extracted. Examples for members of the families
210 *Pseudomonadaceae* and *Enterobacteriaceae* are presented in Supplementary Fig. S1 and S2. In
211 this version of the display the size of sample nodes is related to the cumulative abundance of
212 the taxon and the size of the OTU nodes is related to the cumulative abundance of the OTU in
213 the sub network. Edge thickness gives an estimate of the abundance of a given OTU in each
214 sample, while colours can be used to estimate the relative abundance of food groups in which
215 the selected taxon is found.
216 The node degree distribution for OTU nodes is shown in Supplementary Fig. S3. The
217 distribution fits, albeit with a relatively low R^2 (0.832), a power law distribution with an
218 exponent (γ) of 1.12±0.04. Node degree power law distributions are indicative of a scale-free
219 network (Dunne et al., 2002; Newman et al., 2006). Such networks are widely distributed in
220 all fields (social networks, internet networks, power grids, bibliographic networks) and share
221 several properties. They are usually large and complex, highly connected (large average
222 degree), with a high number of nodes with low degree (in the case of FoodMicrobionet OTUs
223 which are found only in one or few food samples) but with a small numbers of OTUs
224 connected to a large number of samples (i.e. the 'signature' OTUs which make the core
225 microbiota of a given group of food samples). Because FoodMicrobionet 1.0 includes different

food groups, several signature OTUs with high degree are found, and this may affect the fit of the power law distribution.

FoodMicrobionet can also be used to obtain further information on distribution of taxa in different food groups by filtering and recalculation from nodes and edges tables. Information on dominating OTUs can be gathered by plots showing the weighted degree distribution (i.e. how abundant an OTU is in the whole dataset or in a subset) as a function of relative occurrence (i.e. the fractions of samples in which an OTU is found). An example for raw meat is shown in Fig. 3. Further examples for raw milk and mozzarella are shown in Supplementary Fig. S4 and S5. More traditional plots for OTU distribution can also be obtained. An example of the distribution of OTU belonging to different phyla in different food groups is shown in Supplementary Fig. S6.

3.2. Microbial interaction networks.

Microbial interaction networks may help in formulating inferences on the phenomena underlying the structure of food microbial communities, from co-occurrence or co-exclusion patterns due to the occupation of different niches or to selective conditions allowing the growth of a subset of taxa, to relationships such as amensalism, commensalism, symbiosis, etc. The inference of microbial interactions is still affected by pitfalls: the results may be strongly affected by the level of coverage of the microbial community, by the bioinformatics pipelines used with specific options for clustering of the sequences and taxonomic assignment, by the procedures used in normalization and by the methods used to estimate the relationships, etc. However, robust methods have been developed to perform this analysis (Faust et al., 2012). Since OTU abundance tables were available for all datasets included in FoodMicrobionet, we explored co-occurrence/mutual exclusion patterns for all datasets for which a high enough number of samples was available. Statistics for all interaction networks

are shown in Table 3 but a detailed discussion of microbial interactions is beyond the scope of this paper and only two examples are discussed below.

A microbial interaction network for the kefir dataset (Marsh et al., 2013) is shown in Fig. 4. The dataset included milk kefir and grains from different sources. Due to the very simple structure of bacterial communities in kefir and kefir grains only a few interactions were significant. The network has a very low complexity (7 nodes, average degree 3.41 and average path length 2.19), with a clustering coefficient of 0.714, and no fit of the power law for the node degree distribution. The occurrence of *Acetobacter* was negatively related with the occurrence of *Lactobacillales* and that of *Lactobacillus* with *Leuconostoc*, *Lactococcus* and *Streptococcus*. In fact, while *Lactobacillus* dominated the kefir grain microbiota, the latter genera showed a better ability to grow in milk kefir. Members of the family *Lachnospiraceae*, a minor group in the kefir microbiota, also systematically occurred in kefir grains, while they were almost always absent in milk kefir. On the other hand, the co-exclusion relationship with *Acetobacter* was observed in both grains and milk and may reflect conditions for storage and production of kefir.

A very complex interaction network was obtained for the beef dataset (De Filippis et al., 2013). The dataset included swabs from different points of bovine carcasses cuts and beefsteaks obtained thereof, sampled at 0 days and after 7 days of aerobic storage at 4°C, for two different samplings. The full network is shown in Supplementary Fig. S7, while a simplified version, including only the most abundant taxa, is shown in Fig. 5A, together with the interaction network inferred for spoiled beef steak samples (Fig. 5B). The complexity of co-occurrences and mutual exclusions in fresh raw meat mainly reflects the high diversity of bacterial communities (Fig. S7, Fig. 5A). Significant interactions among the most abundant taxa (*Moraxellaceae, Pseudomonadaceae*, *Aerococcaceae, Staphylococcaceae, Flavobacteriaceae, Rhodobacteriaceae* and *Corynebacteriaceae* on carcass swabs and freshly

276 cut beefsteaks; *Pseudomonaceae*, *Listeriaceae*, *Moraxellaceae* and *Enterobacteriaceae* on
277 spoiled steaks) confirm the co-occurrence and mutual exclusion patterns due to different
278 samplings, different cuts, and spoilage described by De Filippis et al. (2013). Spoilage
279 dramatically reduced diversity (De Filippis et al., 2013) and simplified the microbial
280 interaction network (Fig. 5B). The co-occurrence relationship between *Acinetobacter*
281 *guillouiae* (a species occurring at low abundance) and *Enterobacteriaceae* is independent of
282 the sampling and of the cut. On the other hand the mutual exclusion relationship between
283 *Staphylococcus equorum* and *Serratia* is clearly related to the contamination patterns of the
284 beef cuts, with the former species occurring systematically in thick flank cuts and members of
285 the genus *Serratia* occurring in brisket and chuck cuts. The last set of interactions reflects
286 different spoilage environments. In fact, the dominating spoilage organism was *Pseudomonas*
287 in sampling 1 and *Brochothrix* in sampling 2 (De Filippis et al., 2013). *Carnobacterium*,
288 *Acinetobacter johnsonii*, *Chryseobacterium* and members of the class *Actinobacteria* also
289 occurred more frequently in beef steaks from sampling 1.
290 The interaction networks inferred in our study (Table 3) are less complex (sometimes
291 dramatically) than those inferred for environmental bacterial communities (Deng et al., 2012)
292 or for the human microbiome (Faust et al., 2012). In addition, they do not show any fit of the
293 power law for either the node degree distribution or the clustering coefficient/degree
294 relationship, showing that they are neither scale-free nor show a hierarchical structure.
295 However, the average clustering coefficient is often higher than that of random networks with
296 the same size and average degree. In general, interaction networks for fermented or spoiled
297 foods show the lowest complexity, and a high correlation ($r=0.92$) was found between the
298 number of OTUs detected in the dataset and the number of nodes in the interaction network.
299 More complex networks, with >20 nodes were obtained when raw foods (milk, meat) were
300 included in the dataset or when the dataset reflected different environments (milk and

301 cheese, Dolci et al., 2014; Mozzarella produced with different acidification methods, Guidone
302 et al., 2015; raw and spoiled meat, different cuts and samplings, De Filippis et al., 2013). In
303 contrast, Deng et al. (2012) published figures on a wide range of complex bacterial interaction
304 networks from environmental or human sources: the network size ranged from 107 to 254
305 nodes, the node degree distribution showed a good fit of the power law for all networks and
306 the modularity (which measures the occurrence of modules which are strongly
307 interconnected) was significantly higher than that of random networks, while the occurrence
308 of a hierarchical structure was variable. Moreover, the bacterial interaction network of the
309 human microbiome (Faust et al., 2012) included 197 phylotypes with 3005 significant
310 interactions. The network showed a good fit of the power law model for the node degree
311 distribution but did not show a strong hierarchical structure, although the occurrence of body
312 site modules was found. Several factors may contribute to the lower complexity of microbial
313 interaction networks for food. Food microbial communities, and fermented food communities
314 in particular, are dramatically less complex than those found in environmental samples or in
315 the human or animal microbiome, and therefore a lower number of sequences is generally
316 sufficient to obtain a high coverage as it can be predicted by alpha rarefaction analyses
317 (Ercolini, 2013). This may prevent the detection of significant interactions for minor OTUs.
318 Finally, the interactions detected in this study mainly reflect co-occurrence and mutual
319 exclusion patterns in different food environments, and although they in some cases may
320 suggest true positive (commensalism, mutualism) or negative (competition, amensalism)
321 interactions (Gram et al., 2002; Ivey et al., 2013), these should be confirmed in independent
322 experiments.
323
324 *3.3. Future perspectives.*

Both the web visualisation and the full or filtered networks obtained from Gephi, although visually pleasing and informative, are somewhat naïve and should be interpreted with caution. The meta-analyses based on sequencing data published by different laboratories carry some inevitable bias due to differences in data generation and processing. These include possible differences from sample handling through nucleic acid extraction, variable 16S region chosen as target, library purification and preparation, sequencing technology and parameters, sequencing depth / sample coverage (Ercolini, 2013). Furthermore, it is important to underline that the exact bioinformatics path chosen for the analysis can have a strong impact too (May et al., 2014), and will have to be taken into account for a possible standardization of the data handling and usage. In addition, detection of rare OTUs might be affected by biases and reproducibility and repeatibility issues (Benson et al., 2014; Guidone et al., 2015; Pinto and Raskin, 2012; Zhou et al., 2011b). To take this into account it may be advisable to compare different studies at a lower taxonomic resolution and exclude rare OTU from the comparisons. This can be easily done by processing data tables from FoodMicrobionet. An example of a filtered network is shown in Supplementary Fig. S8. In this case, OTUs belonging to the same genus were merged, and the interactive web visualisation is available at http://www.foodmicrobionet.org/fmbn1_0_3gweb/.

Therefore, for a future larger scale meta-analysis it would be advisable to, as a minimum requirement, process the sequences with the same standardized flow in order to limit at least the post-sequencing bias of the analysis. Unfortunately an optimized bioinformatics pipeline is still not available for food microbial communities. Most studies in FoodMicrobionet were carried out using the same sequencing platform and similar or identical bioinformatics pipelines (Table 1) and direct comparisons among studies can be carried out with a good degree of confidence.

With these limitations in mind, the approach used here provides an appealing means for microbiologists and food scientists dealing with food microbial community metadata analysis by (a) providing access to a large set of curated data on the occurrence of different taxa in foods most of which were obtained from studies published in peer reviewed journals, thus facilitating the process of formulating and validating hypotheses on the structure and dynamics of food bacterial communities and writing original articles and reviews; (b) fostering open access to microbial ecology data by making curated nodes and edges tables publicly available; (c) improving our understanding of the ecology of spoilage-associated and beneficial microorganisms; and (d) providing information on the structure of bacterial communities in raw materials, fermented and spoiled foods which can be used for food process development.

While only part of this information is available through the online visualisation, the latter provides a simple interactive interface to explore the microbial ecology of the food environments included in FoodMicrobionet 1.0. Experienced users can import the nodes and edges files provided as supplementary material in a variety of spreadsheet, statistical or network analysis software packages to carry out graphical and statistical analyses or to generate their own networks.

Future plans include (a) expanding the network to other food matrices and food environments; (b) implementing an optimized data analysis pipeline to standardize the treatment of the raw data; and (c) the addition of metadata describing food properties in order to speculate on relevant ecological factors driving microbial interactions and to allow the selection of FoodMicrobionet sub networks with defined range of specific ecological factors. Contributions from other research groups will be welcome. Details on the submission procedure are provided in section 3 of Supplementary Materials.

373 Ultimately, FoodMicrobionet will allow all researchers in the food microbiology to benefit
374 from the significant advances that HTS is providing in this key field of research.
375
376 **References**

377 Bastian, M. H. S., Jacomy, M., 2009. Gephi: An open source software for exploring and
378 manipulating networks. International AAAI Conference on Weblogs and Social
379 Media: San Jose, CA, USA.
380 Benson, A.K., David, J.R.D., Gilbreth, S.E., Smith, G., Nietfeldt, J., Legge, R., Kim, J., Sinha, R.,
381 Duncan, C.E., Ma, J., Singh, I., 2014. Microbial successions are associated with
382 changes in chemical profiles of a model refrigerated fresh pork sausage during an
383 80-day shelf life study. Applied and Environmental Microbiology 80, 5178–5194.
384 Bokulich, N.A, Bamforth, C., Mills, D.A., 2012. Brewhouse-resident microbiota responsible for
385 fermentation of American coolship ale. PLoS One 7, e35507.
386 Bokulich, N. A, Mills, D. A., 2013. Facility-specific 'house' microbiome drives microbial
387 landscapes of artisan cheesemaking plants. Applied and Environmental
388 Microbiology 79, 5214–23.
389 Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer,
390 N., Gonzalez Peña, A., Goodrich, J. K., Gordon, J. I., Huttley G. A., Kelley, S. T., Knights,
391 D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M.,
392 Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko,
393 T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput
394 community sequencing data. Nature Methods 7, 335–336.
395 Chaillou, S., Chaulot-Talmon, A., Caekebeke, H., Cardinal, M., Christieans, S., Denis, C.,
396 Desmonts, M. H., Dousset, X., Feurer, C., Hamon, E., Joffraud, J.-J., La Carbona, S.,
397 Leroi, F., Leroy, S., Lorre, S., Macé, S., Pilet, M.-F., Prévost, H., Rivollier, M., Roux, D.,

398 Talon, R., Zagorec, M., Champomier-Vergès, M.-C., 2015. Origin and ecological
399 selection of core and food-specific bacterial communities associated with meat and
400 seafood spoilage. ISME Journal 9, 1105-1118.
401 Cocolin, L., Alessandria, V., Botta, C., Gorra, R., De Filippis, F., Ercolini, D., Rantsiou, K., 2013.
402 NaOH-debittering induces changes in bacterial ecology during table olives
403 fermentation. PLoS One 8: e69074.
404 Cocolin, L., Ercolini, D., 2015. Zooming into food-associated microbial consortia: a 'cultural'
405 evolution. Current Opinion in Food Science 2, 43–50.
406 De Filippis, F., La Storia, A., Stellato, G., Gatti, M., Ercolini, D., 2014. A selected core microbiome
407 drives the early stages of three popular Italian sheese manufactures. PLoS One 9,
408 e89680.
409 De Filippis, F., La Storia, A., Villani, F., Ercolini, D., 2013. Exploring the sources of bacterial
410 spoilers in beefsteaks by culture-independent high-throughput sequencing. PLoS
411 One 8, e70222.
412 de Menezes, A. B., Prendergast-Miller, M. T., Richardson, A. E., Toscas, P., Farrell, M.,
413 Macdonald, L. M., Baker, G., Wark, T., Thrall, P. H., 2014. Network analysis reveals
414 that bacteria and fungi form modules that correlate independently with soil
415 parameters. Environmental Microbiology 17, 2677–2689.
416 Deng, Y., Jiang, Y.-H., Yang, Y., He, Z., Luo, F., Zhou, J., 2012. Molecular Ecological Network
417 Analyses. BMC Bioinformatics 13, 113.
418 De Pasquale, I., Calasso, M., Mancini, L., Ercolini, D., La Storia, A., De Angelis, M., Di Cagno, R.,
419 Gobbetti, M., 2014a. Causal relationship between microbial ecology dynamics and
420 proteolysis during manufacture and ripening of protected designation of origin
421 (PDO) cheese Canestrato Pugliese. Applied and Environmental Microbiology 80,
422 4085–4094.

423 De Pasquale, I., Di Cagno, R., Buchin, S., De Angelis, M., Gobbetti, M., 2014b. Microbial ecology
424 dynamics reveal a succession in the core microbiota involved in the ripening of
425 Pasta Filata Caciocavallo Pugliese Cheese. Applied and Environmental Microbiology
426 80, 6243–6255.
427 Dolci, P., De Filippis, F., La Storia, A., Ercolini, D., Cocolin, L., 2014. rRNA-based monitoring of
428 the microbiota involved in Fontina PDO cheese production in relation to different
429 stages of cow lactation. International Journal of Food Microbiology 185, 127–135.
430 Dunne, J. A., Williams, R. J., Martinez, N. D., 2002. Food-web structure and network theory: the
431 role of connectance and size. Proceedings of the National Academy of Sciences USA
432 99, 12917–12922.
433 EFSA (2011). The food classification and description system FoodEx 2 (draft-revision 1).
434 Supporting publications 2011:215.
435 http://www.efsa.europa.eu/en/supporting/pub/215e.htm
436 Ercolini, D., 2013. High-throughput sequencing and metagenomics: moving forward in the
437 culture-independent analysis of food microbial ecology. Applied and Environmental
438 Microbiology 79, 3148–3155.
439 Ercolini, D., De Filippis, F., La Storia, A., Iacono, M. (2012). “Remake” by high-throughput
440 sequencing of the microbiota involved in the production of water buffalo mozzarella
441 cheese. Applied and Environmental Microbiology 78, 8142–8145.
442 Ercolini, D., Pontonio, E., De Filippis, F., Minervini, F., La Storia, A., Gobbetti, M., Di Cagno, R.,
443 2013. Microbial ecology dynamics during rye and wheat sourdough preparation.
444 Applied and Environmental Microbiology 79, 7827-7836.
445 Faust, K., Sathirapongsasuti, J. F., Izard, J., Segata, N., Gevers, D., Raes, J., Huttenhower, C., 2012.
446 Microbial co-occurrence relationships in the human microbiome. PLoS
447 Computational Biology 8: e1002606.

448 Faust, K., Raes, J., 2012. Microbial interactions: from networks to models. Nature Reviews
449 Microbiology 10: 538–550.
450 Ferrocino, I., Greppi, A., La Storia, A., Rantsiou, K., Ercolini, D., Cocolin, L. 2016. Impact of
451 nisin-activated packaging on microbiota of beef burgers during storage. Applied and
452 Environmental Microbiology, AEM.03093–15–36. doi: 10.1128/AEM.03093-15
453 Glenn, T.C., 2011. Field Guide to Next Generation DNA Sequencers. Mol. Ecol. Resour. doi:
454 10.1111/j.1755-0998.2011.03024.x
455 Gram, L., Ravn, L., Rasch, M., Bartholin Bruhn, J., Christensen, A. B., Givskov, M.. 2002. Food
456 spoilage—interactions between food spoilage bacteria. International Journal of
457 Food Microbiology 78: 79–97.
458 Greppi, A., Ferrocino, I., La Storia, A., Rantsiou, K., Ercolini, D., Cocolin, L., 2015. Monitoring of
459 the microbiota of fermented sausages by culture independent rRNA-based
460 approaches. International Journal of Food Microbiology 212, 67–75.
461 Guidone, A., Zotta, T., Matera, A., Ricciardi, A., De Filippis, F., Ercolini, D., Parente, E. 2016. The
462 microbiota of high-moisture Mozzarella cheese produced with different acidification
463 methods. International Journal of Food Microbiology 216, 9-17..
464 Hu, Y., 2006. Efficient, high-quality force-directed graph drawing. The Mathematica Journal,
465 10, 37–71.
466 Ivey, M., Massel, M., Phister, T. G. 2013. Microbial interactions in food fermentations. Annual
467 Review of Food Science and Technology 4: 141–62.
468 Kuczynski, J., Liu, Z., Lozupone, C., McDonald, D., Fierer, N., Knight, R., 2010. Microbial
469 community resemblance methods differ in their ability to detect biologically
470 relevant patterns. Nature Methods 7, 813–819.

471 Liu, J., Yang, H., Zhao, M., Zhang, X.-H., 2014. Spatial distribution patterns of benthic microbial
472 communities along the Pearl estuary, China. Systematic and Applied Microbiology
473 37, 578–589.
474 Marsh, A. J., O'Sullivan, O., Hill, C., Ross, R. P., Cotter, P. D., 2013. Sequencing-based analysis of
475 the bacterial and fungal composition of Kefir grains and milks from multiple
476 sources. PLoS One 8, e69371.
477 May, A., Abeln, S., Crielaard, W., Heringa, J., Brandt, B. W. 2014. Unraveling the outcome of 16S
478 rDNA-based taxonomy analysis through mock data and simulations. Bioinformatics
479 30, 1530–38.
480 Mayo, B., Rachid, C. T. C. C., Alegría, A., Leite, A. M. O., Peixoto, R. S., Delgado, S., 2014. Impact
481 of next generation sequencing techniques in food microbiology. Current Genomics
482 15, 293–309.
483 Mendes, L. W., Kuramae, E. E, Navarrete, A. A., van Veen, J. A., Tsai, S. M., 2014. Taxonomical
484 and functional microbial community selection in soybean rhizosphere. ISME Journal
485 8, 1577–1587.
486 Mounier, J., Monnet, C., Vallaeys, T., Arditi, R., Sarthou, A.-S., Hélias, A., Irlinger, F., 2008.
487 Microbial interactions within a cheese microbial community. Applied and Environtal
488 Microbiology 74, 172–81.
489 Muegge, B. D., Kuczynski, J., Knights, D., Clemente, J. C., Gonzalez, A., Fontana, L., Henrissat, B.,
490 Knight, R., Gordon, J. I. , 2011. Diet drives convergence in gut microbiome functions
491 across mammalian phylogeny and within humans. Science 332, 970–974.
492 Newman, M., Barabási, A.-L., Watts, D. J., 2006. The structure and dynamics of networks.
493 Princeton University Press, Princeton.

494 Oakley, B. B., Morales, C. A., Line, J., Berrang, M. E., Meinersmann, R. J., Tillman, G. E., Wise, M.
495 G., Siragusa, G. R., Hiett, K. L., Seal, B. S., 2013. The poultry-associated microbiome:
496 network analysis and farm-to-fork characterizations. Plos ONE 8, e57190.
497 Parente, E., Guidone, A., Matera, A., De Filippis, F., Mauriello, G., Ricciardi, A. 2016. Microbial
498 community dynamics in thermophilic undefined milk starter cultures. International
499 Journal of Food Microbiology 217, 59–67.
500 Pinto, A.J., Raskin, L., 2012. PCR biases distort bacterial and archaeal community structure in
501 pyrosequencing datasets. PLoS One 7, e43093.
502 Rizzello, C. G., Cavoski, I., Turk, J., Ercolini, D., Nionelli, L., Pontonio, E., De Angelis, M., De
503 Filippis, F., Gobbetti, M., Di Cagno, R., 2015. Organic cultivation of *Triticum turgidum*
504 subsp. *durum* is reflected in the flour-sourdough fermentation-bread axis. Applied
505 and Environmental Microbiology 81, 3192–3204.
506 Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R.
507 A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van
508 Horn, D. J., Weber, C. F., 2009. Introducing MOTHUR: open-source, platform-
509 independent, community-supported software for describing and comparing
510 microbial communities. Applied and Environmental Microbiology 75,7537-7541.
511 Zhou, J., Deng, Y., Luo, F., He, Z., Yang, Y., 2011a. Phylogenetic molecular ecological network of
512 soil microbial communities in response to elevated CO_2. mBio 2, e00122–11.
513 Zhou, J., Wu, L., Deng, Y., Zhi, X., Jiang, Y.-H., Tu, Q., Xie, J., Van Nostrand, J.D., He, Z., Yang, Y.,
514 2011b. Reproducibility and quantitation of amplicon sequencing-based detection.
515 ISME Journal 5, 1303–1313.
516

517 **Figure legends**

518

519 **Figure 1.** Flowchart of the development of FoodMicrobionet v 1.0. FoodMicrobionet is a
520 curated database of HTS studies on food bacterial communities which is implemented in
521 Gephi 0.8.2, a network analysis software. The network file can then be used to generate a
522 variety of products for visual and statistical analysis.
523 **Figure 2.** Filtered (only OTU nodes with a cumulative abundance >5% are shown) for raw
524 meat (A) sourdough samples (B) and dairy products and starters (C) extracted from
525 FoodMicrobionet 1.0. Node colour (grey scale for food subgroups for sample nodes or
526 bacterial family of OTU nodes) is used to highlight different sample and OTU nodes. Style
527 features are used to enhance the graph: node size is related to the weighted degree (i.e.
528 cumulative abundances for OTUs) while edge thickness is proportional to the abundance of an
529 OTU in a given sample. Areas of the graph in which samples belonging to a given group are
530 more abundant are enclosed by dashed lines.
531 **Figure 3.** Relative occurrence/ weighted degree scatterplot for OTU nodes in raw meat
532 samples (De Filippis et al., 2013). Only nodes with and weighted degree >1 are shown.
533 Different symbols are used for members of different phyla and the identity of nodes with a
534 weighted degree >5 is shown.
535 **Figure 4.** Microbial interaction network for the kefir dataset (Marsh et al., 2013). Each node
536 represents an OTU. Interactions were evaluated at different taxonomic levels. Only significant
537 interactions are shown ($p<0.0004$; $q<4x10^{-4}$). Edges showing negative interactions (co-
538 exclusion) are coloured red, those for positive interactions in green. The colour of nodes
539 corresponds to the class. The thickness of the edges reflect the level of significance of the
540 supporting evidence for the association (as q-values, $0\text{-}4x10^{-4}$), while the size of the nodes is
541 proportional to their degree.

Figure 5. Microbial interaction network for the beef dataset (De Filippis et al., 2013). A simplified network for all samples (non-spoiled and spoiled, A) and the full interaction network for spoiled beefsteaks (B) are shown. Each node represents an OTU (only low level taxa are shown). Interactions were evaluated at different taxonomic levels. Only significant interactions are shown ($p<0.0004$; $q<4x10^{-4}$). Edges showing negative interactions (co-exclusion) are coloured red, those for positive interactions in green. The colour of nodes corresponds to the class. The thickness of the edges reflect the level of significance of the supporting evidence for the association (as q-values, $0\text{-}4x10^{-4}$), while the size of the nodes is proportional to their degree. Actinobacteriac refers to the class *Actinobacteria*.